

December 18, 1964

To: Dr. L. Cavalli

From: J. Lederberg

1. The field you started: use of demographic data for genetic purposes. Will take a lot of ingenuity to make real use of the actual file, especially a/c limitations of household vs. family units. But we do have access to the 5% sample now with its 10⁷ records. For the next year we will be running test programs on California or some part of it since we want to make only one pass through the whole file.

2. Formal analysis of genetics and molecular biology.

Woodger made a mess of this thirty years ago; it deserves to be taken up again by people who have had some experience with the real world of biological experimentation on the one hand; computers on the other. The organic chemistry algorithms are a primitive example. The first problem is that of syntax. What kinds of assertions "make sense"? DENDRAL does this for the most elementary level of organic chemistry - it can scan a structural assertion to see whether it fits the valence rules; can test whether it is presented in canonical form, and if not, make the necessary transformations; can even generate all valid formulas.

The next step would be to add richer details of real chemistry: the geometrical constraints and generalizations about functional groups.

It would be much more difficult to do this for modern biology, but one should pick out a field - and as Woodger foresaw genetics may be the most apt - and at least try the first step - how to recognize which statements make any sense (whether correct or not). It would certainly be interesting to have a computer generate an orderly census of valid hypotheses of genetics even before considering the next step, of formalizing criteria of correctness.

Walter suggests doing this just in the field of demographic genetics, and he may be right.

Cavalli